Readme File - Group 35

Himanshu	-	2021464
Siddharth	-	2021493
Shivam	-	2021489

October 29, 2023

Schema matching and mapping:

We have follow the materialized view approach :

Schema Matching: Columns from the source tables are matched to create a consistent schema in the view. Each column in the source is matched to its corresponding column in the view (e.g., Title matches Title).

Schema Mapping: Columns that don't exist in a source are mapped to NULL placeholders in the view. Placeholder columns are used to ensure a uniform schema in the view. For example, if a source doesn't have a Description column, a NULL placeholder is used in the view.

ETL/Data exchange/propagation:

Extraction: We have extracted the data through API From our data sources YouTube, Dailymotion and Twitch Platforms through the python code after running that code the data is extracted.

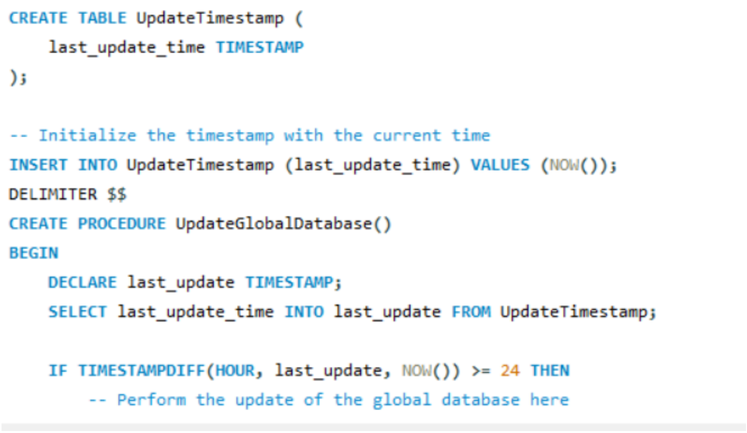
Transformation: We have different units of video durations like in case of twitch , we have duration of videos in minutes while in case of dailymotion and youtube, the duration of videos is in seconds only.

Loading: We have dumped the data through different platforms(YouTube,Dailymotion,Twitch) and dumped this data into three different local schemas. Then we have created a global view for these local schemas.

Incremental View:

Using the Counting Algorithm, we are updating the data from the data sources automatically after every 24 hours or at a fixed time.

Implementation of counting algorithm:



```
CREATE TABLE UpdateTimestamp (
    last_update_time TIMESTAMP
);

-- Initialize the timestamp with the current time
INSERT INTO UpdateTimestamp (last_update_time) VALUES (NOW());
DELIMITER $$
CREATE PROCEDURE UpdateGlobalDatabase()
BEGIN
    DECLARE last_update TIMESTAMP;
    SELECT last_update_time INTO last_update FROM UpdateTimestamp;

    IF TIMESTAMPDIFF(HOUR, last_update, NOW()) >= 24 THEN
        -- Perform the update of the global database here
```

We have updated the total_views of the global database and also total_views from different platforms,total likes in every 24 hours.

Entity Matching: edit distance algorithm (using dynamic programming)/blocking and filtering methods.